Concentration. Loi des grands nombres.

Inégalité de Bienaymé-Tchebychev

Propriété

Soit X une variable aléatoire d'espérance μ et de variance V.
Pour tout réel $\boldsymbol{\delta > 0}$ $\boldsymbol{P(|X-\mu| \geq \delta) \leq \frac{V}{\delta^2}}$

Exemple

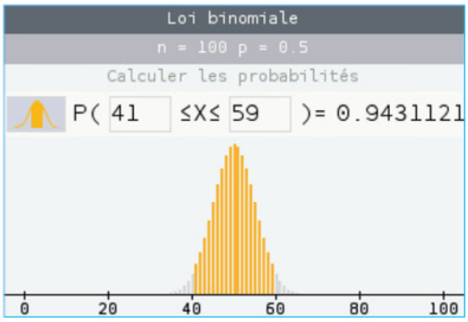
- On lance une pièce équilibrée 100 fois de suite. La variable aléatoire qui compte le nombre de « face » suit une loi binomiale de paramètres $n = 100$ et $p = \frac{1}{2}$

$$E(X) = np = 100 * \frac{1}{2} = 50, V(X) = np(1-p) = 100 * 0{,}5 * 0{,}5 = 25$$

D'après l'inégalité de Bienaymé-Tchebychev, pour tout réel δ

$$P(|X-50| \geq \delta) \leq \frac{25}{\delta^2}$$

- Par exemple pour $\delta = 10$, on a $P(|X-50| \geq 10) \leq 0{,}25$
  La probabilité que l'écart de X à $\mu = 50$ soit supérieure à 10 est inférieure à 0,25
  Or, $P(|X-50| < 10) = 1 - P(|X-50| \geq 10)$
  Donc, $P(|X-50| < 10) \geq 0{,}75$. La probabilité que X prenne une valeur dans l'intervalle $]40\,;60[$ est supérieure à 0,75



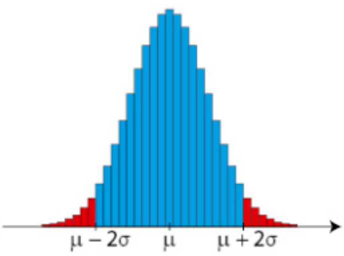
Remarque

Soit X une variable aléatoire d'espérance μ et d'écart type σ.
Nous allons appliquer l'inégalité de Bienaymé-Tchebychev avec des valeurs particulières de δ. En particulier $\delta = k\,\sigma$ où k est un nombre entier naturel. $k \geq 2$

$$P(|X-\mu| \geq k\sigma) \leq \frac{V}{k^2\sigma^2} \leq \frac{\sigma^2}{k^2\sigma^2} \leq \frac{1}{k^2}$$

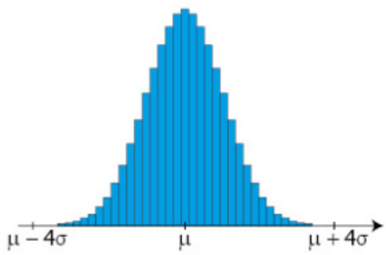
- $k = 2.$

$P(|X-\mu| \geq 2\sigma) \leq \frac{1}{4}$
La probabilité que l'écart de X à μ soit supérieur ou égal à 2σ est inférieure ou égale à 0,25



- $k = 4$

$P(|X-\mu| \geq 4\sigma) \leq \frac{1}{16}$
Obtenir un écart de X à μ supérieur ou égal à 4σ est un événement improbable.



<table>
<tr><th colspan="3">Loi des grands nombres</th></tr>
<tr><td>Contexte</td><td colspan="2">On donne n variables aléatoires ($n \geq 2$) $X_1, X_2, \ldots . X_n$ indépendantes, uniques, suivant une même loi de probabilité d'espérance μ et de variance V
On note $\boldsymbol{M_n} = \frac{X_1+X_2,+\cdots X_n}{n}$ la variable aléatoire moyenne.</td></tr>
<tr><td>Propriété</td><td>Inégalité de concentration</td><td>Pour tout réel $\boldsymbol{\delta > 0}$, $\boldsymbol{P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}}$</td></tr>
<tr><td>Exemple</td><td colspan="2">Dans une société de démarchage par téléphone, on estime que 40% des personnes appelées répondent effectivement. On appelle n personnes, la variable aléatoire X_k donne 1 si la $k-ième$ personne appelée répond et 0 sinon. Les variables aléatoires $X_1, X_2, \ldots X_n$ sont indépendantes, identiques, de même loi d'espérance $\mu = 1 * 0,4 + 0 * 0,6 = 0,4$ et de variance
$$V = 0,4 * (1 - 0,4)^2 + 0,6 * (0 - 0,4)^2 = 0,24$$
Alors pour tout réel $\delta > 0$, $P(|M_n - 0,4| \geq \delta) \leq \frac{0,24}{n\delta^2}$
Par exemple pour $\delta = 0,1$ $P(|M_n - 0,4| \geq 0,1) \leq \frac{0,24}{n0,01} \leq \frac{24}{n}$
Lorsque $n = 1000$, $P(|M_n - 0,4| \geq 0,1) \leq \frac{24}{1000} \leq 0,024$
On dit que l'on obtient pour M_n, une précision de 0,1 avec un risque de 0,024.
Lorsque l'on appelle 1000 personnes, la probabilité que le nombre de personnes qui répondent soit en dehors de l'intervalle $]300; 500[$ est inférieure à 0,024.</td></tr>
<tr><td rowspan="3">Exercice.</td><td colspan="2">Dans une ville moyenne de 20000 habitants, lors d'une consultation portant sur la rénovation du théatre municipal, 75% des personnes consultées ont émis un avis positif.
1. On interroge n personnes. Pour $1 \leq k \leq n$ la variable aléatoire X_k donne 1 si la k-ième personne répond favorablement et 0 sinon. Donner la loi de probabilité de X_k, son espérance et sa variance.
2. On note $M_n = \frac{X_1+X_2,+\cdots X_n}{n}$ la variable aléatoire moyenne.
a. Déterminer une taille n d'échantillon afin d'obtenir pour M_n une précision de 0,05 et un risque de 0,1. C'est-à-dire $P(|M_n - \mu| \geq 0,05) \leq 0,1$
b. De même déterminer une taille n d'échantillon afin d'obtenir pour M_n une précision de 0,01 et un risque de 0,05. Peut-on envisager raisonnablement cette situation ?</td></tr>
<tr><td colspan="2">Solution</td></tr>
<tr><td colspan="2">1. La loi de probabilité de X_k est donnée par $P(X_k = 1) = 0,75$ et $P(X_k = 0) = 0,25$
$E(X_k) = 0,75 * 1 + 0,25 * 0 = 0,75.$ $V(X_k) = E\left(\left(X_k - E(X_k)\right)^2\right)$
$= 0,75(1 - 0,75)^2 + 0,25(0 - 0,25)^2 = 0,1875$
2. D'après l'inégalité de concentration Pour tout réel $\delta > 0$, $P(|M_n - \mu| \geq \delta) \leq \frac{V}{n\delta^2}$
Donc avec $\delta = 0,05$ $P(|M_n - 0,75| \geq 0,05) \leq 0,1$ lorsque $\frac{0,1875}{n0,05^2} \leq 0,1 \Leftrightarrow \frac{75}{n} \leq 0,1 \Leftrightarrow \boldsymbol{n \geq 750}$
De même lorsque $\delta = 0,01$ $P(|M_n - 0,75| \geq 0,01) \leq 0,05$ lorsque $\frac{0,1875}{n0,01^2} \leq 0,05 \Leftrightarrow \frac{1875}{n} \leq 0,05 \Leftrightarrow$ $\boldsymbol{n \geq 37500}$
La ville comptant 20 000 habitants une telle enquête ne peut pas être réalisée. Une précision de 0,01 et un risque de 0,05 ne peuvent donc être envisagés.</td></tr>
<tr><td>Propriété</td><td>Loi des grands nombres.</td><td>Pour tout réel $t > 0, \lim_{n \to +\infty} P(|M_n - \mu| \geq t) = 0$</td></tr>
<tr><td>Remarque</td><td colspan="2">Intuitivement cela signfie que lorsque le nombre de variables aléatoires devient très grand, la probabilité que leur moyenne s'écarte de l'espérance de chacune est très faible</td></tr>
</table>